

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 July 2003 (24.07.2003)

PCT

(10) International Publication Number
WO 03/060766 A1

(51) International Patent Classification⁷: G06F 17/30

LÖNDAHL, Björn [SE/SE]; Glavagatan 23, S-123 71 Farsta (SE).

(21) International Application Number: PCT/SE03/00033

(22) International Filing Date: 13 January 2003 (13.01.2003)

(74) Agents: WIHLSSON, Joakim et al.; Bjerkéns Patentbyrå KB, Östermalmsgatan 58, S-114 50 Stockholm (SE).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0200108-9 16 January 2002 (16.01.2002) SE
60/387,891 13 June 2002 (13.06.2002) US

(81) Designated States (*national*): AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(71) Applicant (*for all designated States except US*): ELUCIDON AB [SE/SE]; Box 45093, S-104 30 Stockholm (SE).

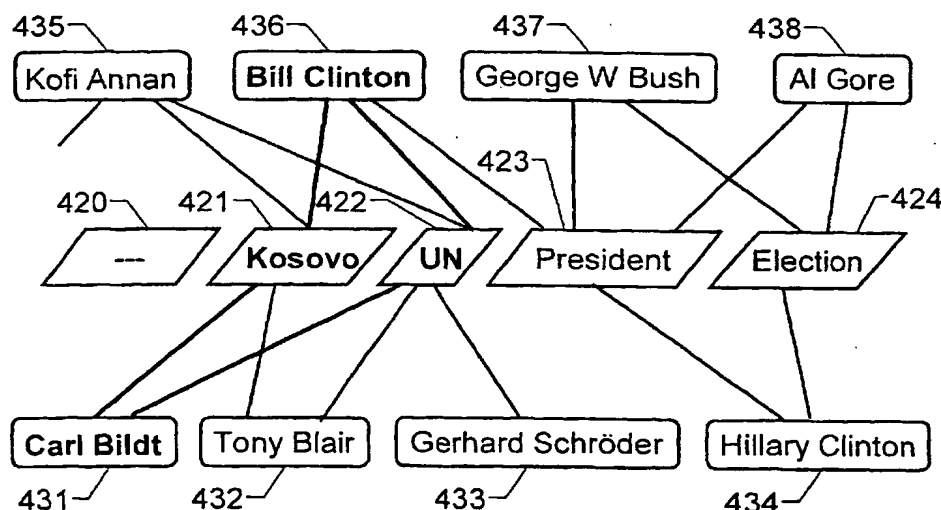
(72) Inventors; and

(75) Inventors/Applicants (*for US only*): LINDH, Per [SE/SE]; Hypoteksvägen 22, S-129 30 Hägersten (SE).

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI,

[Continued on next page]

(54) Title: INFORMATION DATA RETRIEVAL, WHERE THE DATA IS ORGANIZED IN TERMS, DOCUMENTS AND DOCUMENT CORPORA



(57) Abstract: The invention relates to improved solutions for information retrieval, wherein the information is represented by digitized text data. This data is further presumed to be organized in terms (431 - 438), documents and document corpora, where each document contains at least one term (431 - 438) and each document corpus contains at least one document. Based on a concept vector (420 - 424), which conceptually classifies the contents of each document, a term-to-concept vector is generated for each term (431 - 438) in the document corpus. The term-to-concept vector describes a relationship between the term (431) and each of the concept vectors (420 - 424). On basis of the term-to-concept vectors for the document corpus, a term-term matrix is generated which describes a term-to-term relationship between all the terms (431 - 438) in the document corpus. The term-term matrix may then be processed and used for retrieving information from the document corpus, such as the fact that a first term (431) is related to a second term (436).